Accepted Manuscript

A genetic epidemiological study in British adults and older adults shows a high heritability of the combined indicator of vitamin B12 status (cB12) and connects B12 status with utilisation of mitochondrial substrates and energy metabolism.

Anupriya Dalmia, Marie-Joe Dib, Hannah Maude, Dominic J Harrington, Agata Sobczyńska-Malefora, Toby Andrew, Kourosh R Ahmadi

PII: S0955-2863(18)31032-5
DOI: https://doi.org/10.1016/j.jnutbio.2019.04.008
Reference: JNB 8180

To appear in: *The Journal of Nutritional Biochemistry*

Received date: 17 October 2018
Revised date: 25 March 2019
Accepted date: 25 April 2019

Please cite this article as: A. Dalmia, M.-J. Dib, H. Maude, et al., A genetic epidemiological study in British adults and older adults shows a high heritability of the combined indicator of vitamin B12 status (cB12) and connects B12 status with utilisation of mitochondrial substrates and energy metabolism., The Journal of Nutritional Biochemistry, https://doi.org/10.1016/j.jnutbio.2019.04.008

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Title: A genetic epidemiological study in British adults and older adults shows a high heritability of the combined indicator of vitamin B_{12} status (cB_{12}) and connects B_{12} status with utilization of mitochondrial substrates and energy metabolism.

Authors: Anupriya Dalmia[1,2¥], Marie-Joe Dib[1¥], Hannah Maude[2], Dominic J Harrington[3,4], Agata Sobczyńska-Malefora[3,4*], Toby Andrew[2*†] and Kourosh R Ahmadi[1*†]

[1]Department of Nutritional Sciences, School of Biosciences and Medicine, University of Surrey, UK; [2]Department of Genomics of Common Disease, Imperial College, London, UK; [3]The Nutristasis Unit, Viapath, St. Thomas' Hospital, London, UK; [4]King's College London, UK

[¥] These authors contributed equally

[*] These authors contributed equally

[†] Corresponding authors

Correspondence: Department of Nutritional Sciences, School of Biosciences and Medicine, University of Surrey, GU2 7XH, UK. Telephone: +44 (0)1483 684502; E-mail: k.ahmadi@surrey.ac.uk.

Author Contributions

All authors have seen and approved this manuscript. KRA suggested and designed the study. KRA and TA supervised the project. AD, MD, HM, DJH, ASM, TA and KRA were responsible for all statistical/data analyses and interpretation. ASM carried out all the biochemical analyses. DJH and ASM were both instrumental in the initial stages of project development and funding. DJH also wrote part of the manuscript and commented on the final draft of the manuscript. AD was under joint supervision of KRA and TA. MD contributed to writing, discussing and interpreting of the TwinsUK data as well as the clinical data presented in the discussion. All authors helped interpret results and write the report.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Keywords: Vitamin B_{12}; Twin studies; cB_{12}: Heritability: Genetic association study: Mitochondrial function

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

ABSTRACT

Vitamin B_{12} deficiency is common among older adults. However, the most commonly used marker of deficiency, total serum vitamin B_{12} (B_{12}), is not sensitive enough to diagnose true deficiency in a significant proportion of the population. The combined indicator of B_{12} status (cB_{12}), formulated as a composite score of various biomarkers of vitamin B_{12} status (which also accounts for low folate status and age) has been shown to offer a more robust and powerful test to diagnose B_{12} deficiency.

There are no epidemiological studies of cB_{12} variability in older adults. We carried out a twin study to characterise the relative contribution of heritable (h^2) and environmental factors to the observed variability in cB_{12} score in an adult and older adult population (n = 378). Furthermore, we tested for association between variability in cB_{12} and candidate polymorphisms and genes previously associated with B_{12} biomarker levels characterized *in-silico* the mechanism linking the genetic variants and cB_{12} variability.

We found the variability in cB_{12} and its constituents to be highly heritable (h^2= 55%-64%). The SNP *rs291466* in *HIBCH*, previously associated with variation in MMA, was significantly associated with cB_{12} (R^2= 5%|P= 5E-04). Furthermore, variants in *MTRR*, *MMAB*, and *MUT*, underlying inborn errors of B_{12} metabolism, were nominally associated with variation in cB_{12}. Pathway accompanied by expression quantitative trait loci (eQTL) analysis revealed that *HIBCH rs291466* influences the concentration of MMA via the valine degradation pathway.

Our study provides etiological insight into how B_{12} deficiency can manifest into impaired mitochondrial function through perturbations in mitochondrial "fuel" usage.

Conflicts of interest: None.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

INTRODUCTION

Vitamin B_{12}, also known as cobalamin, is an essential micronutrient which in humans acts as a cofactor for two enzymes. In the cytoplasm, B_{12} facilitates re-methylation of homocysteine to methionine by the enzyme methionine synthase (MS), which is encoded by the gene *MTR*. In the mitochondria, B_{12} acts as a cofactor in the conversion of methylmalonyl-CoA to succinyl-CoA by the enzyme methylmalonyl-CoA mutase, encoded by the gene *MUT*, which is thereby important for production of energy [1; 2].

There is a marked variability in B_{12} status and deficiency becomes increasingly common among older adults [3; 4]. The large variability in total serum B_{12} levels among older adults is explained in part by environmental or acquired factors, including inadequate dietary intake, increased malabsorption due to atrophic gastritis, or use of supplements, as well as by heritable factors[5; 6]. Prevention of B_{12} deficiency is of major clinical and public-health importance [7; 8; 9].

Currently, vitamin B_{12} status is mainly assessed through the measurement of total serum vitamin B_{12} (B_{12}) concentrations. Increasingly, however, a combination of other biomarkers e.g. holotranscobalamin (holoTC), methylmalonic acid (MMA) and total homocysteine (tHcy) are used to improve the sensitivity and specificity of assessment in both clinical and epidemiological settings [10]. Notably, the use of these biomarkers individually, as well as the lack of clear cut points to define deficiency, can give rise to contradictory interpretation of vitamin B_{12} status, [11; 12; 13; 14] particularly among the elderly, who are subjected to a myriad of drivers of deficiency (e.g. a higher prevalence of gastritis, renal insufficiency, and use of contraindicated medications, notably metformin, anti-convulsants and proton-pump inhibitors), all of them giving a negative effect on vitamin B_{12} status in older adults[15].

As most of the symptoms associated with cellular functional B_{12} deficiency are due to its metabolic unavailability, it has been suggested that a more useful metric to assess vitamin B_{12} status could be a composite score, cB_{12}, which combines the biochemical measurements of e.g. B_{12}, holoTC with

ACCEPTED MANUSCRIPT

functional biomarkers of B_{12} status – MMA and tHcy while also accounting for low folate status and corrected for age[16]. The combined indicator of vitamin B_{12} status (cB_{12}), previously known as "wellness score", is an index that relates markers of vitamin B_{12} status of the individual to the reference combination at the stipulated age. The reference combination was derived from a large database (N=5,211) following mathematical modelling. The formula which utilizes all four markers of vitamin B_{12} status is expressed as: $cB_{12}=\log_{10}[\text{holoTC x B12/MMA x Hcy}] - [3.79/1+(\text{age}/230)^{2.6}]$. Formulas which utilize two or three markers have also been derived ($2cB_{12}$ or $3cB_{12}$) and offer an improvement on one biomarkers tests. However, the full formula gives a more reliable assessment of B_{12} status[16]. Depending on the cB_{12} value obtained, B_{12} status is classified as: elevated, adequate, low, possible B_{12} deficiency and probable B_{12} deficiency.

Whilst cB_{12} provides a potentially more superior and sensitive marker of B_{12} status, the test is currently still under evaluation and thus it is not commonly used in routine clinical practice. Importantly, from a public health viewpoint, there have been no large-scale genetic epidemiological studies aimed at describing the basis of variability in cB_{12} among adults. We purport that it is of paramount importance to catalogue the principle causes of cB_{12} variability among older adults as a first step in testing its utility in public health programs that aim to redress B_{12} deficiency in older adults.

To fulfil this task and to build on our previous genetic epidemiological study of B_{12}, tHcy, and folate levels in the general population [6; 17], we carried out a classical twin study (n=378 twin pairs) with three broad aims. First, we wanted to characterize the basis of inter-individual variation in cB_{12} among adults and older adults (age 41-82 years old) and apportion the relative contributions from heritable, environmental, and epidemiological factors to variability in cB_{12}. Secondly, guided by previous gene mapping efforts on total B_{12}, homocysteine, MMA, holoTC and folate [6; 17; 18; 19; 20; 21; 22], we set out to validate previous findings and identify new genetic variants responsible for population variation in cB_{12} via genetic association studies. Finally, we aimed to use *in-silico* bioinformatics tools

ACCEPTED MANUSCRIPT

in order to try and functionally characterise the mechanism of action of the implicated genetic variants and pathways in driving B_{12} status determined by cB_{12}.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

MATERIALS AND METHODS

SUBJECTS

Biochemical measurements of serum concentrations of vitamin B_{12}, holoTC, MMA, folate and plasma tHcy along with demographic data (age, BMI, smoking status, alcohol consumption, socioeconomic status and physical activity) were determined for 378 adult female twins from the TwinsUK registry (**Table 1**). All the participants were of British and of Caucasian origin (representative and recruited from across the UK population) [23] from 41 to 82 years old. We excluded from our study any twins who were suffering from a chronic disease (CVD, stroke or diabetes), had liver or kidney dysfunction, received drugs that may interfere with vitamin B_{12} status, or ingested B-complex vitamins or folic acid supplements for up to 3 months before the start of the study.

The present study was approved by the St Thomas' Hospital Research Ethics Committee (EC04/015 TwinsUK). Written informed consent was obtained from all subjects prior to start of the study.

BIOCHEMICAL MEASUREMENTS

Fasting blood samples were collected from each participant and stored on ice. Whole blood was fractionated into serum and plasma within 2 h of collection using standard procedures. Serum/plasma samples were stored at -80 °C until assayed. Serum vitamin B_{12}, holoTC and folate were measured using the Architect 2000 Series analyser (Abbott Diagnostics). The inter-assay CVs% was 5%, 9% and 5% for B_{12}, holoTC and folate, respectively. The laboratory also participates in the UK NEQAS Quality Assurance Scheme for these tests. The cut-offs used for B_{12}, holoTC and folates were 148 pmol/L, 32 pmol/L and 6.8 nmol/L respectively. Serum folate >45.3 nmol/L and serum B_{12} levels >660 pmol/L was considered elevated. Serum MMA was measured by liquid chromatography tandem mass spectrometry (LC-MS/MS). A Gerstel Multi-Purpose Sampler (MPS) coupled directly to LC-MS/MS was used. Intra and inter-assay CV% were <10%. The performance of this test in the DEKS (Aarhus University Hospital) Quality Assurance scheme was satisfactory. Total plasma homocysteine

ACCEPTED MANUSCRIPT

was measured by a rapid high-performance liquid chromatographic assay for total homocysteine[24]. MMA >350 nmol/L and tHcy >15 µmol/L were considered elevated.

PRELIMINARY STATISTICAL ANALYSIS

All preliminary analyses were performed using STATA [25] and R software [26] implemented data summary and linear regression methods. A p-value <0.05 was considered to be nominally significant.

IDENTIFYING COVARIATES OF cB_{12}

We used linear regression to test for the effects of age, BMI, alcohol consumption, smoking status, physical activity, social class, and elevated folate levels (defined liberally as either >30 or >45nmol/L)[27] on cB_{12}, using STATA's iterative robust regression procedure ('*rreg*'), which is robust to outliers and heteroscedasticity. We included elevated folate status as a potential covariate because although cB_{12} has the capacity to be corrected when folate levels are below 10 nmol/L, cB_{12} does not have correction for elevated folate levels, which is traditionally defined as >45 nmol/L. Twin-pair family relatedness was accounted for in these analyses using the 'cluster' option in STATA.

GENETIC MODEL FITTING & HERITABILITY ANALYSIS

To test the twin equal environments assumption (which states that within-twin phenotypic covariance is not confounded by zygosity), we used linear regression ('*reg*', STATA) and test of variance equality ('*sdtest*', STATA) functions to assess equality of means and standard deviations by zygosity for all variables. A classical twin model analysis was used to partition total phenotypic variance (V_P) into genetic (V_G) and environmental (V_E) components using a comparison between monozygotic (MZ) and dizygotic (DZ) pairs. Two standard and complimentary approaches were used to estimate heritability (h^2) of cB_{12}: (i) maximum likelihood variance components analysis using Mx software (www.vcu.edu/mx/), specifically designed for the analysis of twin and family data; and (ii) Defries-Fulker analysis using the '*df*' user implemented function in STATA (Andrew, 2001, available upon request).

ACCEPTED MANUSCRIPT

GENETIC ASSOCIATION STUDIES

We carried out a series of genetic association analyses using genotype data which was already available based on the Illumina 317-Quad array for one half of the study population (n = 159) and the Illumina Human610-Quad Bead Chip for the other half (n = 160). Genotype data was unavailable for 59 twins from the study population.

We applied the following quality control measures across individuals and SNPs to exclude (i) any individuals where the SNP call rate was <95 % and heterozygosity was >37 or <33 % across all SNPs - the expected heterozygosity threshold in a population in the absence of any systematic genotyping or sample-handling errors – and (ii) any SNPs, if they deviated from Hardy–Weinberg equilibrium at $P \leq 10^{-4}$ or had a minor allele frequency of <1%.

Using these data, we carried out three types of genetic association analyses, which were a (i) candidate polymorphism study; (ii) candidate gene study; and (iii) genome wide association study (GWAS). All preliminary genetic association analyses were carried out using PLINK software[28].

<u>Candidate polymorphism study</u>: This study was used to test the hypothesis that previously reported polymorphisms known to be associated with vitamin B_{12}, holoTC, MMA, or tHcy are also associated with variation in cB_{12} among older adults. In order to produce this catalogue of candidate polymorphisms we searched the literature to identify 7 strong candidates that fulfil these criteria. In case of polymorphisms that were not part of the 317k or 610k arrays, we used the SNAP software [29] to identify proxy SNPs that were genotyped on the Illumina Human317-Quad and the Illumina Human610-Quad Bead Chip arrays. The proxy SNPs are chosen based on linkage disequilibrium (LD) threshold (r^2>0.8) and physical distance (<500 kb) between the polymorphism and proxy SNP within the 1000 Genomes dataset, and/or membership in selected commercial genotyping arrays. The list of candidate polymorphisms is shown in **Supplementary Table 1**.

ACCEPTED MANUSCRIPT

Candidate gene study: Using a similar approach used by Andrew et al. (2013), we sought to test the hypothesis that common, less penetrant genetic variants at loci/genes that underlie inborn errors of cobalamin (Cbl) metabolism and transport are also associated with population variability in cB_{12}. The list of candidate genes, their associated defect, and functions of the encoded protein are shown in **Supplementary Table 2**. All array SNPs located within each candidate gene (±10kb from the transcription start and end sites), were included for analysis.

Genome wide association study (GWAS): We carried out a GWAS to identify novel genetic variants that may be associated with population variation in cB_{12} using established pipelines described previously [17; 28].

Genetic association study of mitochondrial DNA variant with cB_{12}: Mitochondrial DNA genotype data for 153 individuals with complete phenotype data were used for this study. A total of 138 SNPs were initially included in the study. All the SNPs with minor allele frequency (MAF) < 0.05 were removed. To control for potential confounding due to population stratification such as different mtDNA haplo-groups, principle component analysis was performed by using the '*pca*' and '*screeplot*' function in STATA for all the SNPs that had complete phenotype data (22 SNPs) for 2,237 TwinsUK samples. The first five principal components from this analysis were used as covariates in the multiple regression analyses for the subsample of 153 individuals with complete cB_{12} phenotype data (data not shown).

Two sets of association analyses were performed, one regressing phenotype (cB_{12} and its biomarkers) upon allele call data and the other regressing phenotype upon quantitative luminous intensity data derived from the Illumina Human610-Quad Bead Chip array for the sample population. The quantitative genotype intensity test was performed in order to check for consistency of results obtained from the allelic association tests and to provide additional confidence to the observed association. To control for multiple testing, the effective number of independent tests[30] was estimated for the mtDNA association study to be a total of 9.9 independent tests based upon the correlation structure between the twenty-seven mtSNPs

ACCEPTED MANUSCRIPT

investigated. For the 6 correlated biomarkers, the effective number of independent tests was 4.5. Hence a conservative Bonferroni corrected threshold α= 0.001 was used for the statistical significance threshold [0.05/(9.9 x 4.5)].

IN-SILICO FOLLOW-UP ANALYSIS OF RESULTS FROM GENETIC ASSOCIATION STUDIES

A particular candidate polymorphism of interest was the *HIBCH* SNP *rs291466*, which is not found on either the Illumina Human317-Quad or Human610-Quad Bead Chip arrays used to genotype samples for this study. Instead we utilised the tagging SNP *rs291458* on both arrays, which is 2.9Kb upstream of *rs291466* and located in intron 1 of *HIBCH* and observed to be in complete linkage disequilibrium (r^2=1) with the SNP of interest.

Linear regression analyses were used to test for association between each biomarker, as well as cB_{12}, with SNPs (candidate polymorphisms and gene regions). In addition to univariate analyses, we also conducted multiple regression analyses to assess whether observed association between the SNPs and MMA were independent of cB_{12}.

Identifying cB_{12} associated SNPs conditionally independent of *rs291466* within *HIBCH:* Multiple regression analyses for each of the 4 biomarkers and cB_{12} on all genotyped SNPs within *HIBCH* ($\pm$10kb) along with the tagging SNP *rs219458* were performed. Similar to the above analysis, an increase in the effect size (adjusted R^2 and significant log-likelihood ratio test) would indicate the role of additional associated *HIBCH* SNPs.

Pathway Analysis: Putative roles of candidate genes *HIBCH* and *MUT* were investigated using the KEGG pathways database and identified to be key genes on the branched chain amino acid (valine, leucine and isoleucine) pathway. To confirm the association between MMA and the valine degradation pathway, association tests using PLINK were conducted. The p-values for univariate regressions of MMA against the SNPs lying within and $\pm$10kb of the candidate genes on the pathway

ACCEPTED MANUSCRIPT

were analyzed. The lists of valine pathway genes tested for genetic association with MMA are listed in **Supplementary Table 3**.

FUNCTIONAL ANNOTATION OF SNP rs291466 in HIBCH

Expression quantitative trait (eQTL) Analysis: We examined the potential functional consequences of the implicated genetic variant *rs291466* by testing the association between proxy markers in complete LD with this SNP and the differential expression of the *HIBCH* gene and other genes encoding enzymes in the valine catabolism pathway using the publicly available expression QTL database GTEx (Genotype-Tissue Expression) [31; 32].

Gene Expression Co-regulation: To test for co-expression between the ten genes of interest (**Supplementary Table 3**) that encode enzymes on the valine catabolism pathway, we downloaded gene expression data for four vitamin B_{12}-relevant tissues from the GEO database [33]. These included liver (GSE41804), whole blood (GSE20489), adipose (GSE15773) and T-cells (GSE26495). Gene expression correlation was estimated using the '*cor*' and '*cor.test*' functions in R (Becker, 1988).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

RESULTS

STUDY POPULATION CHARACTERISTICS

Summary characteristics of the study participants, including demographic and biochemical measurements, were all recorded as continuous variables and are described in **Table 1** for MZ and DZ twin-pairs separately. For each participant, we had data for age, BMI, self-report data on alcohol use, smoking status and employment (used to derive social class using the UK General Registrar Classification), and serum concentrations of B_{12}, holoTC, MMA, folate and plasma tHcy. A total of 80 MZ and 109 DZ female, twin pairs were included in the study.

We excluded data from 7 twins with two or more biomarker measurements indicating B_{12} deficiency: 1) serum B_{12} <148 pmol/L; 2) serum holoTC <32 pmol/L; 3) serum MMA >350 nmol/L; 4) plasma tHcy>15 µmol/L; 5) serum folate <6.8 nmol/L; as well as a cB_{12} cut-off of <-1.5 as stated by Fedosov et al. (2015). In the remaining subjects, we identified: a total of 2 and 3 subjects with folate deficiency (<6.8 nmol/L) and elevated folate status (>45nmol/L), respectively; 3 and 13 subjects with serum B_{12} deficiency and elevated serum B_{12} status (>660 pmol/L), respectively; 3 subjects with serum holoTC <32 pmol/L but normal for all other markers; 8 subjects with elevated plasma tHcy (>15 µmol/L) but normal for all other markers; and 8 subjects with elevated MMA levels (>350 nmol/L) but normal levels of all other markers. In summary, 96-98% of the participants had marker value within the normal range (99% for folate, 96% for B_{12}, 99% for holoTC, 98% for tHcy, and 98% for MMA levels).

PRELIMINARY STATISTICAL ANALYSIS, REPRESENTATIVENESS and TESTING THE EQUAL ENVIRONMENTS ASSUMPTION

Vitamin B_{12}, holoTC, MMA, tHcy, folate and cB_{12} results were normally distributed. The twins (n=378) were generally representative of the females on the TwinsUK database. Monozygotic twins were slightly older than DZ twins in the current data but were comparable for mean weight (**Table 1**),

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

prevalence of smoking, levels of alcohol consumption as well as a number of other demographic and lifestyle variables (data not shown).

There was a minor (~2 years) but significant difference in both means and variance in age by zygosity ($P=0.01$) (**Table 1**). To control for this difference, age was included as a covariate in all subsequent analyses. Our exploratory analysis confirmed that the difference in means and SD by zygosity for all other variables were non-significant ($P>0.05$).

EFFECT OF EPIDEMIOLOGICAL AND DEMOGRAPHIC FACTORS ON cB_{12}

We test for the effect of range of demographic and anthropometric variables on all biomarkers of vitamin B_{12} status and cB_{12}. In this sample of identical and non-identical twins, concentrations of folate, B_{12}, MMA and the cB_{12} score did not show a statistically significant association with any of the anthropometric or demographic variables, including BMI, age, smoking status, alcohol consumption, physical exercise, socio-economic class (**Supplementary Table 4**), or elevated folate levels on cB_{12}. Our data confirmed the positive association between age and tHcy and holoTC.

HERITABILITY OF cB_{12}

Results of the genetic model fitting analysis are shown in **Tables 2**. The intra-class correlation coefficients for the biomarkers of B_{12} status and cB_{12} in MZ and DZ twin pairs are shown in **Supplementary Table 5**. The heritability estimates ranged from 0 (serum folate) to 54% for cB_{12} and 83% for holoTC. These estimates were assessed using both Maximum Likelihood variance components analysis (implemented using Mx) and regression-based DeFries and Fulker (DF) analysis. Our results confirmed a heritability for total serum vitamin B_{12} (h^2= 56%), tHcy (h^2= 64%), and showed for the first time that inter-individual differences in holoTC levels are strongly influenced by heritable factors (h^2= 61%). Concentrations of serum folate and MMA had negligible heritability estimates across all ages and were shown to be predominantly driven by unique environmental (folate), and common and unique environmental factors (MMA) (**Table 2**). However, serum MMA

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

levels did show strong evidence of being heritable (h^2= 33%) among the top age tertile (> 60 years; data not shown).

To assess the impact of age on the heritability of cB_{12} - i.e. to test if the heritable effects become more pronounced with age - we stratified cB_{12} values into 3 age groups and carried out heritability analyses for each group separately. Our results showed a strong trend of increasing genetic and decreasing common environmental influence with age, with the heritability for each age group being 16%, 48% and 63%, for the age groups <56 years, 56-62 years and >62 years, respectively (**Supplementary Table 6**). The most parsimonious and best fitting models for cB_{12} in each age group were ACE/CE (<56 years), AE (56-62 years) and AE (>62 years) for the 3 age groups, respectively.

CANDIDATE SNP-BASED ASSOCIATION STUDY

We searched the literature to identify polymorphisms that have been shown to be significantly associated with serum levels of B_{12}, MMA, or holoTC levels, focusing only on validated/replicated SNPs [(21)]. A total of 7 polymorphisms (MAF≥ 1%) in 7 genes, including *ABCD4, ACSF3, FUT2, MUT, HIBCH, TCN1,* and *CUBN* were identified and included in our candidate polymorphism study (**Supplementary Table 1**). We tested for associations between each candidate polymorphism and/or neighbouring SNPs in high linkage disequilibrium (LD) (in cases where data on this polymorphism was unavailable) and cB_{12} levels using the PLINK software. No "tag" SNPs could be identified for the candidate polymorphism *rs526934* in gene *TCN1* and we were unable to identify adequate proxy SNPs for *TCN1* using SNAP.

We found that variability in cB_{12} was significantly associated with proxy markers in complete LD with the candidate SNP *rs291466* in the gene *HIBCH* (P= 5 x 10^{-4}) and possibly associated with the candidate polymorphism *rs1801222* in gene *CUBN* (P= 0.05) (**Table 3)**.

ASSOCIATION OF GENES UNDERLYING INBORN ERRORS OF COBALAMIN METABOLISM WITH cB_{12}

ACCEPTED MANUSCRIPT

A total of 42 SNPs (MAF≥ 1%) from 8 genes - including the intragenic regions ± 10kb of the start and stop codons that harbour mutations that cause inborn errors of B_{12} metabolism and transport were included in the candidate gene analysis (**Supplementary Table 2**). We present meta-analysed p-values based on analysis from the two different genotyping arrays. Our analyses show that polymorphisms in genes *MMAB* (P= 0.02), *MTR* (P= 0.04) and *MUT* (P= 0.006) to be nominally associated with cB_{12} (α= 0.05) although none passed the Bonferroni corrected significance threshold for "n" multiple comparisons (n=43, p=0.05/43 ≈ 0.001) (**Table 4**).

GENOME-WIDE ASSOCIATION STUDY RESULTS

Acknowledging the power constraints of our cohort in terms of size, we undertook a preliminary GWAS to formally test and assess if there are any loci with moderate to high penetrance variants that underpin the heritability of cB_{12}. Our analysis of genotyped SNPs from the 317K and 610K arrays did not identify any genome-wide significant loci at the significance level of $P \leq 10^{-7}$. A total of 21 SNPs were identified that were nominally associated with cB_{12} ($4 \times 10^{-7} \leq p \leq 10^{-5}$) (**Supplementary Table 7 and Supplementary Figure 1**).

PATHWAY AND IN-SILICO FUNCTIONAL ANNOTATION OF SIGNIFICANT LOCI

Univariate regression results show that the candidate polymorphism *rs291466* in *HIBCH* accounts for approximately 12% of the variation in MMA and 5% of the variation in cB_{12} (**Supplementary Table 8**). Linear regression residuals analysis[34] - using the residuals for cB_{12} regressed upon MMA - confirmed that the candidate polymorphism association with cB_{12} (R^2=0.05, p= 5E-5) appeared to be entirely driven by the SNP association with MMA, with tagSNP *rs291458* no longer associated with cB_{12} when conditioned upon MMA (R^2=1E-5, p=0.86). In addition, the results from the multiple regression of MMA against the *HIBCH* SNPs *rs7590991, rs6753459*, *rs2582762* and *rs291467*, along with tagging SNP *rs291458*, showed that despite the inclusion of these four additional SNPs, with each strongly associated with MMA in univariate analyses, the coefficient of determination value only increased from R^2=0.12 to R^2=0.14, since these SNPs are in linkage disequilibria (associated) with one another

ACCEPTED MANUSCRIPT

(i.e. the four named *HIBCH* SNPs were collectively associated with MMA, independent of tagSNP *rs291458*, with R^2=0.02, χ^2_4=17.1, p=2E-3).

Candidate Pathway Study. *HIBCH* lies on the valine degradation pathway and codes for a hydrolase enzyme that catalyzes the conversion of 3-hydroxy-isobutyrate coenzyme-A to its ketone form (2-methyl-3-oxopropanoate). This implicates the valine degradation pathway as the functional molecular mechanism that underpins the observed association between *HIBCH* and MMA variation. The observed association between the *HIBCH* tag SNP *rs291458* (in complete LD with *rs291466*) and cB_{12} provides new evidence that the valine degradation pathway also impacts upon functional B_{12} status.

In-silico Gene Expression and eQTL analysis. *HIBCH* expression levels are consistently correlated with *MUT, MCEE* and *ABAT* from the branched chain amino acid (valine, leucine and isoleucine) pathway and *PCCA* (transcript 1) in all four tissues examined (liver, whole blood, T cells and adipose). A high degree of correlation was observed in whole blood, in comparison to the other tissues (liver, T cells and adipose - **Supplementary Table 9 and Supplementary Figure 2**). We observed that *rs291466* regulates the expression of *HIBCH* in >30 tissues, i.e. this variant is a *cis* eQTL for *HIBCH* (for whole-blood, p = 9.3×10^{-11}, effect size = 0.27). These results are replicated in all the tissues examined, in which *rs291466* acts as an eQTL for *HIBCH.* Furthermore, our *in-silico* analyses show that *rs291466* is also a *trans*-eQTL for a number of other genes encoding enzymes on the valine degradation pathway, including *ALDH, MUT, ACSF3, HADH and ABAT* (**data not shown**).

Mitochondrial DNA variant association study of with cB_{12}. Mitochondria contain a unique, circular genome that is 16.7 kb long and encodes 13 proteins involved in energy metabolism, 2 ribosomal RNAs (rRNAs) and 22 transfer RNAs (tRNAs) involved in mitochondrial protein translation[35]. Mitochondrial genetic variants (MtSNPs) have been implicated in several common diseases and

ACCEPTED MANUSCRIPT

metabolic disorders[36]. Given the strong link between vitamin B_{12} status and mitochondrial function, we tested whether common genetic variations (MAF ≥ 0.05) within the mitochondrial genome are associated with cB_{12} variation. A total of 27 common SNPs were included in the analysis (**Supplementary Table 10**). Seven SNPs lay in the D-loop (the non-coding control region of the mitochondrial genome) and the remaining 22 SNPs were coding variants. The first 5 components for the 27 mtDNA SNPs accounted for a cumulative 82% of the total genotypic variance in the mitochondrial DNA data using over 2,000 TwinsUK individuals with complete Illumina 610 array data including mtDNA SNPs (**data not shown**). These 5 eigenvectors, therefore, describe a series of uncorrelated linear combinations that contain most of the common variance across the molecule.

The allelic association tests showed strong evidence of association between (1) the mtSNP, *3198T>C* (residing in gene mt-RNR2 encoding a 16s rRNA), and serum concentrations of B_{12} (P= 8E-05) and with cB_{12} (nominal significance, P= 4E-03). The same nominally significant results were obtained when B_{12} (P= 0.03) and cB_{12} (P= 0.004) were assessed for association using mtDNA genotype intensity data; and (2) mtSNP, *1191T>C* (that lies in the gene mt-RNR1 encoding a 12s rRNA), and tHcy levels (P= 6E-05) and nominally associated with holoTC (P= 0.03) concentrations (**Table 5**).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

DISCUSSION

We conducted the first genetic epidemiological study of cB_{12}, the combined indicator of vitamin B_{12} status, in older adults. We show that variations in cB_{12}, as well as tHcy and for the first time holoTC, to be highly heritable, thereby motivating gene mapping efforts for cB_{12} and its constituents. Previous GWAS have identified a number of genetic variants associated with serum concentrations of B_{12} or MMA [18; 19; 20; 21; 37]. In agreement with the findings from these studies, our candidate polymorphism and gene studies show cB_{12} to be nominally associated with SNPs in genes previously linked with biomarkers of B_{12} status (*HIBCH*), rare inborn errors of cobalamin metabolism/transport (*MUT, MMAB* and *MTRR*), as well as mitochondrial function.

In our candidate polymorphism study, we observed significant association between a proxy marker for the SNP *rs291466* in *HIBCH* and cB_{12} with it accounting for 5% of the variation in cB_{12}. The *rs291466* SNP has previously been shown to account for ~10% variation in MMA levels [18]. Our results also showed a 12% of the variation in MMA concentrations is accounted for by the SNP *rs291466*. A log-likelihood ratio test confirmed that the association between the proxy SNP for *rs291466* is associated with cB_{12} and is primarily driven by association with MMA. There could be three plausible reasons for the elevated estimate of 12% of MMA variance, obtained in our study. It is possible that the previous study has underestimated the results, or that the results are the same but reflect sample variation, or that there are additional variants in the *HIBCH* gene region that may account for MMA variation. We tested the third hypothesis by performing a multiple regression analysis with all the genotyped SNPs in the candidate gene region for *HIBCH* along with our candidate polymorphism *rs291466* and found that the effect size did indeed modestly (R^2=0.02), but significantly (p=0.002) increase with the inclusion of 4 additional SNPs from the *HIBCH* candidate gene region. This indicates that, in addition to the previously reported polymorphism *rs291466* associated with MMA[18] and depending upon the population observed, there may be additional

ACCEPTED MANUSCRIPT

polymorphisms in the *HIBCH* gene region that can be identified, that are less strongly, but independently associated with MMA.

A candidate pathway study revealed a potential mechanism for the SNP *rs291466* impact on MMA via the valine degradation pathway, in which *HIBCH* catalyses a step upstream of MMA production. Further *in-silico* analysis revealed that *HIBCH* is co-expressed with at least three other genes of the valine degradation pathway, including the B_{12}-dependent methyl-malonyl-coA mutase (*MUT*) which converts MMA to succinyl-coA, in addition to *ABAT* and *MCEE*. *HIBCH* is not significantly co-expressed with a previously implicated valine catabolism pathway gene, *ACSF3* [(18)] for adipose and liver tissue. The latter observation indicates that the influences of the two candidate polymorphisms (*ACSF3 rs1054747* and *HIBCH rs291466*) on MMA concentration are likely to be independent of each other. The SNP rs291466 is an eQTL that causes an increase in *HIBCH* transcription, possibly causing a correlative change in the co-expressed gene levels (for *ABAT, MCEE* and *MUT*) as a part of a common pathway. The genes *ABAT* and *MUT* are both *trans*-regulated by the eQTL *rs291466* and are co-expressed with *HIBCH*, consistent with the idea that these genes may be co-regulated by a shared molecular mechanism. MMA was also associated with the SNPs in the candidate gene regions for *HADH, ABAT* and *ALDH*, with all of these genes catalyzing steps leading to MMA production as part of valine catabolism.

It is important to note that as stated by Molloy et al [(18)], the variation in MMA as a result of the *HIBCH* SNP *rs291466* or other genes in valine degradation pathway has, in fact, no relation to vitamin B_{12} status itself although an additional increase in MMA can exacerbate the consequences of B_{12} deficiency. The same is true for the genetic variants of all enzymes related to the steady state levels of tHcy and MMA, because a low activity of, let us say, mutated methionine synthase can be observed even at a high B_{12} level. The cB_{12} index was designed to counteract irrelevant fluctuations in the individual markers, i.e. those not caused by low/high B_{12}. Based on our observations, the

ACCEPTED MANUSCRIPT

combined index seems to be capable of "buffering" the changes in MMA and tHcy not related to the "true" B_{12} levels.

In summary, five noteworthy conclusions can be drawn from our study. Firstly, variation in cB_{12} is heritable and driven by common genetic variation. Secondly, our results show that the heritable effects on cB_{12} become more pronounced with age and the effect of environmental factors diminish with age, with this study strongly indicating that the observed increase in population cB_{12} variance is driven by a corresponding increase in population genetic variance with age. Hence, the causes of vitamin B_{12} deficiency in old age may be largely due genetic factors for some sections of the aging population and in principle, genetic risk factors may help identify who these people are. Thirdly, a previously implicated genetic variant in *HIBCH* can be identified as accounting for 5% of the variation in cB_{12} as a function of its correlation with MMA, likely due to its eQTL effects on enzymes of the valine degradation pathway in which MMA is metabolized. Fourthly, it appears that, depending upon the population observed, there may be common genetic *HIBCH* variants, in addition to *rs291466*, that can be identified that account for a modest degree of MMA variation. Finally, the genetic variant *rs291466* is a strong candidate to be included in a panel of markers designed to diagnose vitamin B_{12} deficiency. We verified that the genetic variant *rs291466* (T>C) is a gain-of-function variant that increases the rate of *HIBCH* transcription.

Strengths and limitations of the study

Our study is the first to investigate the variability of cB_{12} in older adults. In addition to having a considerably large sample size of twins, and having data available on four B_{12} biomarkers, and folate status, we were able to run our analysis considering single biomarkers, combination of them in cB_{12}, and the capacity to compute $4cB_{12}$, the strongest proxy of B_{12} status. Furthermore, we have used the most advanced genetic association tests and bioinformatics tools and were able to demonstrate that

ACCEPTED MANUSCRIPT

our findings are implicated in the valine degradation pathway, and may have clinical and/or public health relevance, thus providing a stepping stone in the advancement of precision nutrition.

Our study also has a number of limitations which warrant further discussion. A primary limitation is that the sample population was entirely female and hence, sex-related differences in cB_{12} and its biomarkers could not be studies or accounted for. From a genetic analysis perspective, our study was generally underpowered for the high number of tests carried out. Also, using genotype data from two separate arrays and the fact that meta-analysis could only be conducted for the SNPs common to both the Illumina 317-Quad and the Illumina Human610-Quad Bead Chip arrays, low coverage for some of the candidate gene regions due to low LD (*MMAA*, *MMACHC*, *MMAADHC* and *MTRR)* and inadequate representation of candidate polymorphisms/genes that have previously been shown to be associated with B_{12} biomarkers, including for example *FUT6 rs3760775, rs78060698* (total B_{12}), and *TCN2 rs1131603* (holoTC), or unknown, were technical limitations associated with the study. So, a replication cohort, preferably consisting of males and females, would be critical to confirm some of our preliminary findings. Another limitation is that the study was limited only to a Caucasian population with adequate B_{12} status limiting our capacity to detect more associations because of not having individuals with B_{12} deficiency. Thus, the portability of our findings across ethnicities and other population age groups is yet to be investigated. Lastly, the interaction between B_{12}, folate and genetics was not accounted for in our analysis and may have influenced its outcome. It would be interesting to investigate this in future studies.

Finally, we used a Bonferroni correction which is a conservative method for controlling for the Type II error rate when undertaking multiple tests that are correlated with one another. To remedy this and in order to calculate an adjusted threshold of statistical significance, we first estimated the total number of effectively independent tests performed for the polymorphism and candidate gene association tests, which for this study corresponded to a less conservative significance threshold of $P = 0.001$ (see Methods).

ACCEPTED MANUSCRIPT

Scope and implications of the study

We have estimated that *HIBCH rs291466* accounts for 5% of cB_{12} variation but the mechanism of its action requires further research. Functional analysis including other substrates in the valine degradation pathway could potentially explain the mechanism driving the co-expression between the genes of this pathway. Studies involving patients with vitamin B_{12} deficiency might help determine whether the genetic variant *HIBCH rs291466* is implicated in specific conditions or diseases associated with vitamin B_{12} deficiency, including, vascular, neurodegenerative and metabolic diseases.

An intriguing prospect of our findings is the potential etiological insight into how vitamin B_{12} deficiency can reduce mitochondrial function through perturbations in mitochondrial “fuel” usage and energy metabolism[38]. The conversion of methylmalonyl-coA, produced by the degradation of valine, isoleucine and odd-chain fatty acids, to succinyl-coA is one of many anaplerotic reactions which generate intermediates of the TCA cycle. B_{12} deficiency may therefore have interesting consequences with respect to the rate of incorporation of these metabolites as sources of energy production. A hypothesise to be tested in future studies would be that B_{12} deficiency, driven by environmental (ageing, drug use) and/or genetic factors, impacts upon mitochondrial fuel partitioning, and subsequently energy metabolism, and associated with long-term health outcomes. In this regard, our findings also potentially implicate the mechanism underpinning the observed interaction between increased risk of B_{12} deficiency and metformin use, a drug known to affect mitochondrial function [39; 40] and fuel partitioning[41]. Shen et al. (2017) [42; 43] recently showed that loss-of-function (LoF) variants in gene *CLYBL* leads to defects in mitochondrial-B_{12} metabolism through the accumulation of an intermediate in the C5-dicarboxylate metabolic pathway, itaconate-derived itaconyl-CoA, which acts as a potent inhibitor of the mitochondrial B_{12}-dependent methylmalonyl-CoA mutase (MUT). We postulate that the effect of metformin on B_{12} status can mimic the *CLYBL* defect. Metformin might disturb the mitochondrial C5 pathway and neutralise vitamin B_{12} leading to elevations in MMA levels. Inhibition of MUT will, in addition, perturb

ACCEPTED MANUSCRIPT

mitochondria's ability to use branched chain amino acids (BCAAs) and odd chain fatty acids and affect mitochondrial fuel utilisation and function. Future studies should aim to investigate the combined effects of *CLYBL* LoF-mimicking effect of metformin on B_{12} neutralisation and perturbations in mitochondrial substrate utilisation caused by common genetic variation in genes encoding key intermediates of the BCAA oxidation pathway.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Acknowledgments

KRA would like to acknowledge support from the BBSRC DTP grant. TA would like to acknowledge support from an MRC New Investigator Research Grant (91993). TwinsUK NIHR BRC BioResource is funded by the Wellcome Trust, the European Community's Seventh Framework Programme (FP7/2007-2013), and the National Institute for Health Research (NIHR) BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. We would like to thank all the study participants as well as staff from the TwinsUK study.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

References

1. Froese DS, Gravel RA (2010) Genetic disorders of vitamin B(1)(2) metabolism: eight complementation groups--eight genes. *Expert reviews in molecular medicine* **12**, e37.

2. Janos Zempleni JWS, Jesse F. Gregory III, Patrick J. Stover (2017) Handbook of Vitamins: CRC.

3. Clarke R, Grimley Evans J, Schneede J *et al.* (2004) Vitamin B12 and folate deficiency in later life. *Age and ageing* **33**, 34-41.

4. Loikas S, Koskinen P, Irjala K *et al.* (2007) Vitamin B12 deficiency in the aged: a population-based study. *Age and ageing* **36**, 177-183.

5. van Asselt DZ, de Groot LC, van Staveren WA *et al.* (1998) Role of cobalamin intake and atrophic gastritis in mild cobalamin deficiency in older Dutch subjects. *The American journal of clinical nutrition* **68**, 328-334.

6. Andrew T, Gill R, Gillham-Nasenya I *et al.* (2013) Unravelling the basis of variability in cobalamin levels in the general population. *The British journal of nutrition*, 1-8.

7. Smith AD, Refsum H (2012) Do we need to reconsider the desirable blood level of vitamin B12? *Journal of internal medicine* **271**, 179-182.

8. Carmel R (2012) Subclinical cobalamin deficiency. *Current opinion in gastroenterology* **28**, 151-158.

9. Baik HW, Russell RM (1999) Vitamin B12 deficiency in the elderly. *Annual review of nutrition* **19**, 357-377.

10. Sobczynska-Malefora A, Gorska R, Pelisser M *et al.* (2014) An audit of holotranscobalamin ("Active" B12) and methylmalonic acid assays for the assessment of vitamin B12 status: application in a mixed patient population. *Clinical biochemistry* **47**, 82-86.

11. Green R (2011) Indicators for assessing folate and vitamin B-12 status and for monitoring the efficacy of intervention strategies. *Am J Clin Nutr* **94**, 666S-672S.

12. Carmel R, Green R, Rosenblatt DS *et al.* (2003) Update on cobalamin, folate, and homocysteine. *Hematology Am Soc Hematol Educ Program*, 62-81.

ACCEPTED MANUSCRIPT

13. Obeid R, Schorr H, Eckert R *et al.* (2004) Vitamin B12 status in the elderly as judged by available biochemical markers. *Clin Chem* **50**, 238-241.

14. Nexo E, Hoffmann-Lucke E (2011) Holotranscobalamin, a marker of vitamin B-12 status: analytical aspects and clinical utility. *The American journal of clinical nutrition* **94**, 359S-365S.

15. Hunt A, Harrington D, Robinson S (2014) Vitamin B12 deficiency. *BMJ* **349**, g5226.

16. Fedosov SN, Brito A, Miller JW *et al.* (2015) Combined indicator of vitamin B12 status: modification for missing biomarkers and folate status and recommendations for revised cut-points. *Clinical chemistry and laboratory medicine : CCLM / FESCC* **53**, 1215-1225.

17. van Meurs JB, Pare G, Schwartz SM *et al.* (2013) Common genetic loci influencing plasma homocysteine concentrations and their effect on risk of coronary artery disease. *The American journal of clinical nutrition* **98**, 668-676.

18. Molloy AM, Pangilinan F, Mills JL *et al.* (2016) A Common Polymorphism in HIBCH Influences Methylmalonic Acid Concentrations in Blood Independently of Cobalamin. *Am J Hum Genet* **98**, 869-882.

19. Hazra A, Kraft P, Lazarus R *et al.* (2009) Genome-wide significant predictors of metabolites in the one-carbon metabolism pathway. *Human molecular genetics* **18**, 4677-4687.

20. Hazra A, Kraft P, Selhub J *et al.* (2008) Common variants of FUT2 are associated with plasma vitamin B12 levels. *Nature genetics* **40**, 1160-1162.

21. Grarup N, Sulem P, Sandholt CH *et al.* (2013) Genetic architecture of vitamin B12 and folate levels uncovered applying deeply sequenced large datasets. *PLoS genetics* **9**, e1003530.

22. Lin X, Lu D, Gao Y *et al.* (2012) Genome-wide association study identifies novel loci associated with serum level of vitamin B12 in Chinese men. *Human molecular genetics* **21**, 2610-2617.

23. Moayyeri A, Hammond CJ, Valdes AM *et al.* (2013) Cohort Profile: TwinsUK and healthy ageing twin study. *International journal of epidemiology* **42**, 76-85.

ACCEPTED MANUSCRIPT

24. Ubbink JB, Hayward Vermaak WJ, Bissbort S (1991) Rapid high-performance liquid chromatographic assay for total homocysteine levels in human serum. *Journal of chromatography* **565**, 441-446.

25. LP S (2015) Stata Statistical Software, 14 ed.

26. Team RC (2018) R: A language and environment for statistical computing. [RFfS Computing, editor]. Vienna, Austria.

27. Sobczynska-Malefora A, Harrington DJ (2018) Laboratory assessment of folate (vitamin B9) status. *Journal of clinical pathology* **71**, 949-956.

28. Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559-575.

29. Johnson AD, Handsaker RE, Pulit SL *et al.* (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24**, 2938-2939.

30. Camp NJ, Farnham JM (2001) Correcting for multiple analyses in genomewide linkage studies. *Annals of human genetics* **65**, 577-582.

31. Lonsdale J, Thomas J, Salvatore M *et al.* (2013) The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585.

32. Consortium GT (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-660.

33. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207-210.

34. Hills C (1993) *Statistical Models in Epidemiology*: Oxford University Press

35. Wallace DC (1999) Mitochondrial diseases in man and mouse. *Science* **283**, 1482-1488.

36. Wallace DC (2005) A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annual review of genetics* **39**, 359-407.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

37. Tanaka T, Scheet P, Giusti B *et al.* (2009) Genome-wide association study of vitamin B6, vitamin B12, folate, and homocysteine blood concentrations. *American journal of human genetics* **84**, 477-482.

38. Gao X, Lee K, Reid MA *et al.* (2018) Serine Availability Influences Mitochondrial Dynamics and Function through Lipid Metabolism. *Cell reports* **22**, 3507-3520.

39. de Jager J, Kooy A, Lehert P *et al.* (2010) Long term treatment with metformin in patients with type 2 diabetes and risk of vitamin B-12 deficiency: randomised placebo controlled trial. *BMJ* **340**, c2181.

40. Luciano-Mateo F, Hernandez-Aguilera A, Cabre N *et al.* (2017) Nutrients in Energy and One-Carbon Metabolism: Learning from Metformin Users. *Nutrients* **9**.

41. Liu X, Romero IL, Litchfield LM *et al.* (2016) Metformin Targets Central Carbon Metabolism and Reveals Mitochondrial Requirements in Human Cancers. *Cell metabolism* **24**, 728-739.

42. Shen H, Campanello GC, Flicker D *et al.* (2017) The Human Knockout Gene CLYBL Connects Itaconate to Vitamin B12. *Cell* **171**, 771-782 e711.

43. Reid MA, Paik J, Locasale JW (2017) A Missing Link to Vitamin B12 Metabolism. *Cell* **171**, 736-737.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 1. Characteristics of the study population. Means (± standard deviations) and inter-quartile ranges (IQR) are given for a total of 368 female twins. The p-value of the Wilcox-Whitney-Mann test reflects whether there were any significant differences between the zygosity groups ($p < 0.05$*).

	MZ		DZ		p-value
	Mean	IQR	Mean	IQR	
n	160 (80 pairs)		218 (109 pairs)		
Age (years)	60.7 ± 7.9	54.0 - 67.0	58.6 ± 6.80	53.0 - 63.0	**0.01***
Height (cm)	162.9 ± 5.3	158.0 - 167.0	162.8 ± 5.3	159.0 - 167.0	0.77
Weight (kg)	71.2 ± 11.5	63.8 - 75.2	68.6 ± 12.56	59.7 - 73.9	0.16
BMI (kg/m^2)	26.4 ± 3.5	23.9 - 28.4	25.2 ± 3.67	22.3 -27.8	0.09
Folate (nmol/l)	19.8 ± 9.5	12.6 - 24.7	21.3 ± 9.77	13.8 - 27.5	0.14
Vitamin B_{12} (pmol/l)	378.3 ± 167.1	275.6 - 458.0	379.5 ± 162.7	276.7 - 464.1	0.71
holoTC (pmol/l)	99.6 ± 68.2	57.5 - 111.2	87.2 ± 49.67	58.0 - 103.4	0.28
MMA (nmol/l)	162.9 ± 75.2	105.0 - 202.0	170.0 ± 84.21	120.5 - 196.0	0.63
tHcy (µmol/l)	12.4 ± 3.4	9.8 - 14.3	12.5 ± 3.04	10.2 - 14.2	0.66
cB_{12}	0.54 ± 0.52	0.22 – 0.85	0.46 ± 0.49	0.13 – 0.81	0.18

MZ = monozygotic; DZ = dizygotic; holoTC = holotranscobalamin; MMA = methylmalonic acid; tHcy = total homocysteine, cB12 = composite score

ACCEP

Table 2. Heritability of 1-CM biomarkers and the composite score cB_{12} concentrations. Results show a strong heritability for vitamin B_{12}, holoTC, tHcy and cB_{12}. Folate and MMA are predominantly influenced by common and unique environmental factors.

	Model	$\chi 2$	df	$\Delta\chi 2$	Δdf	*P*	A	C/D	E	AIC
Folate	ACE	1.45	3	-	-	-	0.11 (0.00 - 0.54)	0.3 (0.00 - 0.49)	0.58 (0.43 - 0.74)	-4.54
	AE	3.93	4	2.48	1	0.28	0.46 (0.30 - 0.59)	-	0.54 (0.41 - 0.69)	-4.06
	CE	**1.67**	**4**	**0.22**	**1**	**0.93**	**-**	**0.38 (0.24 - 0.49)**	**0.62 (0.50 - 0.75)**	**-6.33**
Vitamin B_{12}	ACE	15.49	3	-	-	-	0.46 (0.03 - 0.67)	0.09 (0.00 - 0.43)	0.45 (0.32 - 0.61)	9.49
	AE	**15.74**	**4**	**0.25**	**1**	**0.93**	**0.56 (0.42 - 0.67)**	**-**	**0.44 (0.32 - 0.58)**	**11.86**
	CE	19.86	4	4.37	1	<0.05	-	0.42 (0.29 - 0.53)	0.57 (0.46 - 0.70)	11.86
holoTC	ACE	7.38	3	-	-	-	0.50 (0.09 - 0.70)	0.11 (0.00 - 0.44)	0.39 (0.29 - 0.53)	1.38
	AE	**7.65**	**4**	**0.27**	**1**	**0.92**	**0.61 (0.48 - 0.70)**	**-**	**0.39 (0.29 - 0.51)**	**-0.35**
	CE	13.25	4	5.87	1	<0.05	-	0.49 (0.37 - 0.59)	0.51 (0.41 - 0.63)	5.25
MMA	ACE	5.12	3	-	-	-	0.06 (0.00 - 0.54)	0.42 (0.00 - 0.59)	0.51 (0.37 - 0.66)	-0.78
	AE	9.15	4	4.03	1	<0.05	0.53 (0.37 - 0.65)	-	0.47 (0.35 - 0.63)	1.15
	CE	**5.28**	**4**	**0.16**	**1**	**0.95**	**-**	**0.47 (0.33 - 0.59)**	**0.52 (0.41 - 0.66)**	**-2.72**
tHcy	ACE	6.59	3	-	-	-	0.64 (0.35 - 0.74)	0.00 (0.00 - 0.23)	0.35 (0.25 - 0.48)	0.59
	AE	**6.59**	**4**	**0**	**1**	**0**	**0.64 (0.51 - 0.74)**	**-**	**0.35 (0.25 - 0.48)**	**-1.4**
	CE	20.25	4	13.66	1	1.59E-06	-	0.44 (0.31 - 0.55)	0.55 (0.44 - 0.68)	12.25
cB_{12}	ACE	4.5	3	-	-	-	0.39 (0.00 - 0.69)	0.14 (0.00 - 0.47)	0.47 (0.34 - 0.65)	-1.4
	AE	**5.2**	**4**	**0.7**	**1**	**0.78**	**0.55 (0.40 - 0.67)**	**-**	**0.45 (0.33 - 0.60)**	**-2.8**
	CE	7.6	4	3.1	1	0.18	-	0.41 (0.28 - 0.52)	0.59 (0.47 - 0.71)	-0.4

ACCEPTED MANUSCRIPT

$\Delta\chi 2$ = Change in $\chi 2$ fit statistic or likelihood ratio test (LRT); A, additive polygenic variance component; C, shared familial or common environment; E, unique environmental effects (including measurement error) that are specific to the individual. * Three models are presented, each with best-fit model statistics and standardised variance component estimates for A, C/D and E that add up to one for each model. The most parsimonious and best-fit model is highlighted, with a heritability estimate (95 % CI) and the remaining vriance attributed to environmental factors specific to the individual. Model-fit comparisons are based on a LRT, where the model-fit contribution for each parameter (A, D and C) is assessed by individually fixing the value of each to zero to contrast the full model (ACE) with a nested sub-model (AE and CE). If the LRT is significant (here assessed by the change in $\chi 2$ with one df), the variance component contributes to the model fit and cannot be dropped. Common environment (C) and dominant genetic (D) effects are confounded using twin heritability models, so these components cannot be estimated in the same model.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 3. Results from the candidate polymorphism association study of cB_{12}.

Gene	Candidate SNP	Chr	Proxy SNP	MAF	Coordinate	Distance	r^2	D'	Major	Minor	BETA	SE	R^2	P
HIBCH	*rs291466*	2	*rs2664253*	0.44	190,856,803	2,994	1.0	1.0	G	A	0.15	0.04	0.05	**5.E-04**
			rs291458	0.49	190,889,726	35,917	1.0	1.0	A	C	0.15	0.04	0.05	**5.E-04**
			rs291460	0.48	190,889,830	2,890	1.0	1.0	G	A	0.15	0.04	0.05	**5.E-04**
			rs291429	0.43	190,906,186	13,466	0.9	1.0	G	A	0.19	0.06	0.06	**1.E-03**
ACSF3	*rs1054747*	16	*rs729201*	0.05	87,768,284	19,937	1.0	1.0	A	G	-0.14	0.20	0.00	0.50
FUT2	*rs492602*	19	*rs602662*	0.41	53,898,797	568	0.8	1.0	A	G	-0.05	0.07	0.00	0.49
			rs485186	0.41	53,899,018	789	0.8	1.0	G	A	-0.05	0.07	0.00	0.49
			rs504963	0.41	53,900,677	2,448	0.8	1.0	A	G	-0.05	0.07	0.00	0.49
			rs676388	0.43	53,903,781	5,552	0.8	1.0	C	T	-0.04	0.06	0.00	0.58
MUT	*rs9473558*	6	*rs9473541*	0.38	49,461,316	59,076	0.9	1.0	C	A	-0.06	0.07	0.01	0.32
			rs2501976	0.38	49,478,943	41,449	0.9	1.0	C	T	-0.07	0.07	0.01	0.30
			rs2501979	0.38	49,484,883	35,509	0.9	1.0	T	C	-0.07	0.07	0.01	0.30
			rs6458690	0.37	49,519,578	814	1.0	1.0	A	G	-0.06	0.07	0.01	0.36
			rs4267943	0.37	49,547,764	27,372	1.0	1.0	G	A	-0.06	0.07	0.01	0.34
ABCD4	*rs12896084*	14	*rs3213499*	0.29	73,826,292	29,999	0.8	1.0	A	G	0.07	0.06	0.01	0.27
			rs3742801	0.27	73,828,759	27,532	1.0	1.0	C	T	0.07	0.06	0.01	0.29
			rs4148078	0.27	73,829,054	27,237	1.0	1.0	G	T	0.07	0.06	0.01	0.29
			rs4148077	0.27	73,829,230	27,061	1.0	1.0	C	T	0.07	0.06	0.01	0.29
			rs3742800	0.28	73,831,337	24,954	0.9	1.0	A	G	0.07	0.06	0.01	0.29
			rs2301347	0.27	73,835,646	20,645	1.0	1.0	C	A	0.07	0.06	0.01	0.29
			rs2301346	0.26	73,835,833	20,458	0.9	1.0	T	C	0.03	0.07	0.00	0.70

ACCEPTED MANUSCRIPT

			rs2301345	0.27	73,836,105	20,186	1.0	1.0	A	G	0.07	0.06	0.01	0.29
CUBN	*rs1801222*	10	*rs12261966*	0.32	17,183,006	13,151	0.6	0.9	C	T	-0.11	0.06	0.02	**6.E-02**
TCN1	*rs526934*	*11*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*

Kb=kilo base pairs.

¶ The number of genotyped single nucleotide polymorphisms (SNPs) for each gene region

▼The degree to which these markers are pair-wise correlated (r2) with all common SNPs documented by the HapMap project for these regions.

RefSNP ID= reference SNP ID number; c2= Chi-squared statistic.

¶Significant association with Cbl levels were based on a conservative Bonferroni corrected P≤ 0.001.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 4. Results from the candidate gene association study of cB_{12}. Significant associations was based on a conservative Bonferroni correction, P≤ 10-3.

Gene	Defect	Location	RefSNP ID	P
MMAA	cblA	4q31	*rs1497125*	0.5
			rs4835011	0.33
			rs4835012	0.24
			rs4835014	0.88
			rs6831899	0.17
MMAB	cblB	12q24	*rs2241206*	2.20E-02
			rs7134594	0.62
			rs3759387	0.55
			rs7957619	0.59
MMACHC	cblC	1p34	*rs1494813*	0.34
			rs487174	0.11
MMADHC	cblD	2q23	*rs4667423*	0.4
			rs7580915	0.36
MTR	cblE	5p15	*rs1046014*	5.80E-02
			rs326120	0.89
			rs1532268	0.18
			rs6555501	7.00E-02
			rs162031	0.54
			rs162036	0.25
			rs162040	0.36
			rs162048	0.71

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

			rs3776455	0.99
			rs10380	9.30E-02
			rs7715062	0.11
LMBRD1	cblF	6q13	*rs6912049*	0.49
			rs6455338	0.59
			rs7742015	0.61
			rs3799105	0.71
			rs3778241	0.8
			rs9294851	0.6
			rs991974	0.49
			rs1457498	0.96
			rs4304126	0.57
MTRR	cblG	1q43	*rs2275568*	5.00E-02
			rs3768142	0.6
			rs1805087	4.00E-02
			rs3820571	0.75
MUT	mut	6p21	*rs9381784*	**1.00E-02**
			rs6458687	**6.00E-03**
			rs6458690	0.5
			rs6458692	0.15
			rs4267943	0.52
Abbreviations: RefSNP ID, reference SNP ID number; cbl, cobalamin.				

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 5. Results of allelic association tests between common mitochondrial variants and static, functional biomarkers of vitamin B_{12} status including cB_{12}.

BIOMARKER	Variant	P	N
Vitamin B_{12}	*MitoT3198C*	**8E-05***	151
Folate	*MitoA3481G*	2.00E-02	151
	MitoA10551G	2.00E-02	151
	MitoA14234G	4.00E-02	151
	MitoT14799C	2.60E-03	147
holoTC	*MitoT1191C*	3.00E-02	152
	MitoT9699C	4.00E-02	152
MMA	*MitoG16130A*	7.00E-03	144
tHcy	*MitoG228A*	3.00E-02	148
	MitoT491C	4.00E-02	148
	MitoT1191C	**6E-05***	148
	MitoG16130A	3.00E-03	143
cB_{12}	*MitoT3198C*	4.00E-03	152

* mtSNPs that passed Bonferroni correction (a = 0.001);

N, number of observations

RIPT